



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Genome-wide regional heritability mapping identifies a locus within the TOX2 gene associated with Major Depressive Disorder

Citation for published version:

Zeng, Y, Navarro, P, Shirali, M, Howard, D, Adams, M, Hall, L, Clarke, T-K, Thomson, P, Smith, BH, Murray, A, Padmanabhan, S, Hayward, C, Boutin, T, MacIntyre, D, Lewis, CM, Wray, NR, Mehta, D, Penninx, B, Milaneschi, Y, Baune, BT, Air, T, Hottenga, J-J, Mbarek, H, Castelao, E, Pistis, G, Schulze, TG, Streit, F, Forstner, AJ, Byrne, EM, Martin, NG, Breen, G, Müller-Myhsok, B, Lucae, S, Kloiber, S, Domenici, E, Deary, I, Porteous, D, Haley, C & McIntosh, A 2017, 'Genome-wide regional heritability mapping identifies a locus within the TOX2 gene associated with Major Depressive Disorder', *Biological Psychiatry*, vol. 82, no. 5, pp. 312-321. <https://doi.org/10.1016/j.biopsych.2016.12.012>

Digital Object Identifier (DOI):

[10.1016/j.biopsych.2016.12.012](https://doi.org/10.1016/j.biopsych.2016.12.012)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Biological Psychiatry

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

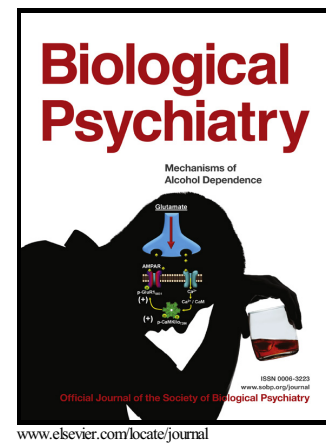
Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Genome-wide regional heritability mapping identifies a locus within the TOX2 gene associated with Major Depressive DisorderRegional Heritability Mapping for MDD

Yanni Zeng, Pau Navarro, Masoud Shirali, David M. Howard, Mark J. Adams, Lynsey S. Hall, Toni-Kim Clarke, Pippa A. Thomson, Blair H. Smith, Alison Murray, Sandosh Padmanabhan, Caroline Hayward, Thibaud Boutin, Donald J. MacIntyre, Cathryn M. Lewis, Naomi R Wray, Divya Mehta, Brenda Wjh Penninx, Yuri Milaneschi, Bernhard T. Baune, Tracy Air, Jouke-Jan Hottenga, Hamdi Mbarek, Enrique Castelao, Giorgio Pistis, Thomas G Schulze, Fabian Streit, Andreas J. Forstner, Enda M Byrne, Nicholas G Martin, Gerome Breen, Bertram Müller-Myhsok, Susanne Lucae, Stefan Kloiber, Enrico Domenici, Ian J Deary, David J Porteous, Chris S. Haley, Andrew M. McIntosh



PII: S0006-3223(16)33113-4
DOI: <http://dx.doi.org/10.1016/j.biopsych.2016.12.012>
Reference: BPS13069

To appear in: *Biological Psychiatry*

Cite this article as: Yanni Zeng, Pau Navarro, Masoud Shirali, David M. Howard, Mark J. Adams, Lynsey S. Hall, Toni-Kim Clarke, Pippa A. Thomson, Blair H. Smith, Alison Murray, Sandosh Padmanabhan, Caroline Hayward, Thibaud Boutin, Donald J. MacIntyre, Cathryn M. Lewis, Naomi R Wray, Divya Mehta, Brenda Wjh Penninx, Yuri Milaneschi, Bernhard T. Baune, Tracy Air, Jouke-Jan Hottenga, Hamdi Mbarek, Enrique Castelao, Giorgio Pistis, Thomas G Schulze, Fabian Streit, Andreas J. Forstner, Enda M Byrne, Nicholas G Martin, Gerome Breen, Bertram Müller-Myhsok, Susanne Lucae, Stefan Kloiber, Enrico Domenici, Ian J Deary, David J Porteous, Chris S. Haley and Andrew M. McIntosh, Genome-wide regional heritability mapping identifies a locus within the TOX2 gene associated with Major Depressive DisorderRegional Heritability Mapping for MDD, *Biological Psychiatry*, <http://dx.doi.org/10.1016/j.biopsych.2016.12.012>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and

review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Genome-wide regional heritability mapping identifies a locus within the *TOX2* gene associated with Major Depressive Disorder

Short title: Regional Heritability Mapping for MDD

Yanni Zeng^{1*}, Pau Navarro², Masoud Shirali^{2,23}, David M. Howard¹, Mark J. Adams¹, Lynsey S. Hall¹, Toni-Kim Clarke¹, Pippa A. Thomson^{3,4}, Blair H. Smith^{5,22}, Alison Murray⁶, Sandosh Padmanabhan^{4,7}, Caroline Hayward⁴, Thibaud Boutin⁴, Donald J. MacIntyre¹, Cathryn M. Lewis⁸, Naomi R Wray⁹, Divya Mehta⁹, Brenda Wjh Penninx¹⁰, Yuri Milaneschi¹⁰, Bernhard T. Baune¹¹, Tracy Air¹¹, Jouke-Jan Hottenga¹², Hamdi Mbarek¹², Enrique Castelao¹³, Giorgio Pistis¹³, Thomas G Schulze^{14,15,16}, Fabian Streit¹⁷, Andreas J. Forstner¹⁸, Enda M Byrne⁹, Nicholas G Martin¹⁹, Gerome Breen⁸, Bertram Müller-Myhsok²⁰, Susanne Lucae²⁰, Stefan Kloiber²⁰, Enrico Domenici²¹, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium[^], Ian J Deary^{3,22,23}, David J Porteous^{3,4,23}, Chris S. Haley^{2,24} and Andrew M. McIntosh^{1,3,23}

Affiliations:

¹ Division of Psychiatry, University of Edinburgh, Edinburgh, UK

² MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

³ Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, United Kingdom

⁴ Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh

⁵ Division of Population Health Sciences, University of Dundee, Dundee, UK

⁶ Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

⁷ Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

⁸ MRC social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

⁹ Queensland Brain Institute, University of Queensland, St Lucia, Queensland, Australia

¹⁰ Department of Psychiatry, VU University Medical Center, Amsterdam, Netherlands

¹¹ Discipline of Psychiatry, University of Adelaide, Adelaide Australia

¹² Department of Biological Psychology, VU University, Amsterdam, the Netherlands

¹³ Department of Psychiatry, CHUV, Lausanne, Switzerland

¹⁴ Institute of Psychiatric Phenomics and Genomics, Ludwig-Maximilians-University, Munich, Germany

¹⁵ Department of Psychiatry and Psychotherapy, University Medical Center, Georg-August-University, Göttingen, Germany

¹⁶ Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg, Germany

¹⁷ Department of Genetic Epidemiology in Psychiatry, Medical Faculty Mannheim, Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany

¹⁸ Institute of Human Genetics, University of Bonn, Bonn, Germany; Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany

¹⁹ School of Psychology, University of Queensland, St Lucia, Queensland, Australia

²⁰ Max Planck Institute of Psychiatry, Germany Munich Cluster for Systems Neurology (SyNergy), Germany.

²¹ Laboratory of Neurogenomic Biomarkers, Centre for Integrative Biology, University of Trento, Italy.

²² Department of Psychology, University of Edinburgh, UK

²³ Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

²⁴ The Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, UK

* Corresponding author

Yanni Zeng

Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK EH10 5HF, T: +44 (131) 537 6687

E: Y.Zeng-6@sms.ed.ac.uk

[^] Group author

Abstract

Background: Major Depressive Disorder (MDD) is the second largest cause of global disease burden. It has an estimated heritability of 37% but published genome-wide association studies have so far identified few risk loci. Haplotype-block-based regional heritability mapping (HRHM) estimates the localized genetic variance explained by common variants within haplotype blocks, integrating the effects of multiple variants, and maybe more powerful for identifying MDD-associated genomic region.

Methods: We applied HRHM to GS:SFHS, a large family and population based Scottish cohort (N=19,896). Single-SNP and haplotype-based association tests were used to localize the association signal within the regions identified by HRHM. Functional prediction was used to investigate the effect of MDD-associated SNPs within the regions.

Results: A haplotype block across a 24kb region within the *TOX2* gene reached genome-wide significance in HRHM. Single-SNP and haplotype-based association tests demonstrated that five out of nine genotyped SNPs and two haplotypes within this block were significantly associated with MDD. The expression of *TOX2* and a brain-specific lncRNA RP1-269M15.3 in frontal cortex and Nucleus accumbens basal ganglia, respectively, were significantly regulated by MDD-associated SNPs within this region. Both the regional heritability and single SNP-associations within this block were replicated in the UK-Ireland group of the most recent release of the Psychiatric Genomics consortium (PGC2-MDD). The SNP-association was also replicated in a depressive symptom sample that shares some individuals with PGC2-MDD.

Conclusion: This study highlights the value of HRHM for MDD and provides an important target within *TOX2* for further functional studies.

Key Words: Regional heritability; HRHM; *TOX2*; MDD; haplotype block; Genome-wide analyses

Introduction

Major Depressive Disorder (MDD) is ranked as the second leading contributor to the global disease burden in terms of the years lived with disability(1). The narrow sense heritability of MDD has been estimated to be approximately 37% by twin studies(2), suggesting a substantial contribution from genetic factors. In efforts to identify specific genetic risk factors for MDD, family-based linkage studies have identified several significant peaks in certain families, but the findings have been inconsistent(3). GWAS studies of unrelated participants have successfully identified hundreds of loci associated with other psychiatric disorders(4), but for MDD, only four genome-wide significant and replicable loci have been identified by two large GWAS studies, one on a refined MDD phenotype for Chinese Women and one on self-report-based depression using less intensive phenotyping in a much larger European sample(5-7).

Several factors may be responsible for the comparatively sparse GWAS results in MDD. First, MDD is likely to have a highly polygenic genetic architecture where the disease risk is conferred by many causal variants of small effect(8,9). Combined with the high prevalence of MDD(10) and the possible incomplete LD between genotyped SNPs and causal SNPs, single-SNP based genome-wide association tests may have insufficient power to detect individual causal variants(11). Second, clinical heterogeneity has been shown in MDD between populations(6,12), and this may lead to difficulties in identifying causal variants across cohorts(13). Whilst GWAS sample sizes for MDD are increasing and efforts to refine MDD phenotype are in progress(5,7), alternative methodologies for detecting the signal arising from causal variants within and across families may also be productive.

Regional heritability mapping (RHM) is a method used to identify small genomic regions accounting for a significant proportion of the phenotypic variance in a trait of interest(14). In contrast to single-SNP based tests, RHM integrates effects from multiple SNPs by utilizing a regional genetic relationship matrix estimated from SNPs within a region. The matrix is constructed for each region defined by a sliding window across the genome, and is then used to estimate the variance explained by the variants within the region in a linear mixed model(14). The major advantage of RHM is that the regional genetic relationship matrices not only tag the effect of genotyped variants, but also measure the effect of un-genotyped and rare variants, including those associated with the SNPs but with individual effects too small to be detected by GWAS(14,15). Previous studies have shown that RHM has greater power to detect rare variants and multiple alleles in regions where GWASs provided null findings(15-17). In 2014, Shirali *et.al* developed a haplotype-block-based heritability mapping (HRHM) method as an improved version of RHM. HRHM utilizes haplotype blocks as the unit of mapping therefore the identified blocks have less complex local LD structures(18).

In this study, we applied HRHM to a homogenous sample of approximately 20K Scottish participants containing both closely- and distantly-related subjects with genome-wide genotyping data and a standardised structured clinical MDD diagnosis(19). We sought to identify genomic regions conferring risk for MDD, which were then further explored using single-SNP- and haplotype- based association tests. We then examined the functional effects of the MDD-associated SNPs within the identified block. Finally, replication analyses were performed in independent samples

for both the regional heritability and SNP association results.

Methods and Materials

The Tayside Research Ethics Committee (reference 05/S1401/89) provided ethical approval for the study. Participants all gave written consent, after having an opportunity to discuss the project, and before any data or samples were collected.

Datasets

Discovery sample: Generation Scotland: The Scottish Family Health Study (GS:SFHS) contains 21,387 subjects ($N_{\text{male}}=8,772$, $N_{\text{female}}=12,615$; $\text{Age}_{\text{mean}}=47.2(\text{SD}=15.1)$), who were recruited from the registers of collaborating general practices in Glasgow, Tayside, Ayrshire, Arran and Northeast regions of Scotland. At least one first-degree relative aged 18 or over was required to be identified for each participant(19,20). A structured clinical interview was used for the diagnosis of lifetime DSM-IV mood disorders (SCID)(21,22). Details of MDD diagnosis, genotyping, quality control (QC) and imputation methods are described in the Supplement. In total 561,125 genotyped and 8,642,105 post-imputation autosomal SNPs passed QC criteria were available for 19,896 participants (2,659 MDD cases and 17,237 controls) for subsequent analyses.

Replication sample 1: UK Biobank

Data used in this study were provided as part of UK Biobank project reference #4844.

Details for genotyping, quality control, imputation and phenotyping are described in the Supplement. In brief, genotyping data was available for 152,729 UK Biobank participants recruited in United kingdom(23). The probable MDD phenotype was created based on the putative MDD definition established in Smith et al(2013) using responses to a touchscreen questionnaire (UK Biobank 2011b)(24), from self-report information, and from inpatient records via linkage to hospital episode data (Supplement). After quality control and removing subjects who were in both GS:SFHS and UK Biobank datasets, and one of each pair of close relatives (relatedness > 0.05) of GS:SFHS participants or the remained UK Biobank participants, 1,198,327 SNPs for 24,015 subjects with putative MDD phenotype available (8,143 cases and 15,872 controls) remained in downstream analyses.

Replication sample2: PGC Major Depression Dataset (PGC2-MDD)

The Psychiatric Genomics Consortium provided individual genotypes (best guess) of imputed SNPs for participants from 22 cohorts in PGC2-MDD(Table s1). All cases met DSM-IV criteria for life MDD, the majority of them were ascertained clinically. Most control samples were screened and participants with lifetime MDD were removed(Table s1). Details for genotyping, quality control, imputation and phenotyping are described in the Supplement. After quality control and removing subjects overlapped with GS:SFHS and UK Biobank dataset, 32,554 subjects of European ancestry (13,261 cases and 19,293 controls) were used in downstream analysis. Consistent with earlier work(25,26), we grouped the 22 cohorts into 7 groups based on the country of ancestor information for regional heritability analysis(Table s1).

Replication sample3: Depressive symptom datasets (DS) (this sample contains overlapping individuals with replication sample 1 and 2)

Okbay *et.al*(2016) carried out a GWAS meta-analysis ($N = 180,866$) on three samples using depressive symptoms as the trait of interest (27). The ascertained MDD diagnosis information was available for two samples (PGC1-MDD($N_{\text{cases}} = 9,240$, $N_{\text{controls}} = 9,519$) and the Resource for Genetic Epidemiology Research on Aging (GERA, $N_{\text{cases}} = 7,231$, $N_{\text{controls}} = 49,316$))(27). For the third sample (UK Biobank ($N = 105,739$)), a continuous phenotype measuring the severity of depressive symptom had been created and used in the meta-analysis(27). Whilst this sample overlapped with the PGC2-MDD and UK Biobank samples, it provided results based upon a non-diagnostic quantitative measure of depressive symptoms and involved another large cohort, GERA(27).

Genome-wide haplotype-block-based Regional Heritability Mapping (HRHM)

Regional heritability mapping (RHM) is a method for detecting localized genomic regions where genetic variants contribute significantly to the variation of phenotype of interest(14). As an improved version of RHM, HRHM divides the genome into haplotype-blocks, based on the recombination hotspots in the genome(18). Details of HRHM are described in the Supplement. In brief, in GS:SFHS the genotyped SNPs were mapped to 49,637 haplotype-blocks across the genome and the regional heritability was estimated and tested for each of the haplotype-blocks. A standard ‘two-GRM’ model incorporates two genomic relationship matrices (GRM); a regional

genomic relationship matrix (rGRM) estimated from SNPs in the haplotype block and a complement genomic relationship matrix (cGRM) estimated from all SNPs that are not included in the haplotype block. These GRMs were jointly fitted as random effects in LMM. Covariates fitted as fixed effects include age, age², sex, 20 principal components. A Log likelihood ratio test (LRT) is applied to test the significance of random effect represented in rGRM by comparing a model with both cGRM and an rGRM fitted against a model including the cGRM but without an rGRM fitted. The genome-wide significance threshold for P values from LRT is 1.01×10^{-6} ($N_{Bonferroni}=49,637$). This two-GRM model, while providing an unbiased estimate of regional heritability, was highly computationally demanding. To improve the calculation efficiency, a pre-adjustment strategy was applied in the genome-wide HRHM (Supplement). For haplotype-blocks that exceeded the genome-wide significant threshold, we re-tested the block using the two-GRM model to provide an accurate estimation of regional heritability in the target block. All the analyses were performed in REACTA(14,28). According to the GCTA-GREML Power Calculator, this study is well-powered for the GREML-based SNP heritability analysis (99.88%)(29).

Localized association tests for the significant haplotype block identified by HRHM in GS:SFHS

HRHM identified a significant block chr20:42555671-42579473, we performed a series of association tests to localize the association signals within this block in GS:SFHS.

- 1) Single-SNP-based association test for common SNPs within the identified

haplotype-block.

Association tests were performed on genotyped and imputed common SNPs located in the significant haplotype-block chr20:42555671-42579473 using GCTA-MLMA (linear mixed model based association analysis)(30). The SNP effect was tested as a fixed effect, other covariates included age, age², sex and 20 PCs. To prevent the estimates of SNP effects from being confounded by the polygenic component and family structure, cGRM and cGRM_{kin} were fitted simultaneously as random effects in the model(31). cGRM (complement-snp-set GRM) was the genomic relationship created matrix using all of the genotyped SNPs excluding the SNPs in the hit block; cGRM_{kin} was the kinship relationship matrix (representing pedigree-associated genetic variation). cGRM_{kin} was created by setting elements in cGRM that were less than or equal to 0.05 to 0(31). The estimated fixed effect (on the linear scale) was transformed to logit and liability scale using Taylor series approximation(32). Bonferroni multiple-testing-correction was performed for the P values for each SNP.

2) Single-haplotype-based association test

Single-haplotype-based association tests were performed for the common haplotypes (Frequency \geq 0.01) derived from the nine genotyped common SNPs located in the significant haplotype-block chr20:42555671-42579473 using GCTA-MLMA(30) for the full dataset and an unrelated dataset, and using famLBL(33) for a subset consisting of case-parent trios in GS:SFHS. Details of single-haplotype-based association test are described in the Supplement.

Functional effects of MDD-associated-SNPs in the significant block

The significant haplotype-block chr20:42555671-42579473 is located in the intron region and a proportion of an adjacent exon of gene *TOX2*. To investigate the potential functional effects from variants within this block, we imputed the 9 genotyped SNPs within this block to 53 common SNPs based on HRC reference, all of them are non-coding SNPs. We performed single-SNP-based association test for each of them with MDD using GCTA-MLMA (the same method for genotyped SNPs). This identified 38 imputed SNPs significantly associated with MDD. We then examined the functional role of the 38 SNPs using the following functional annotation tools and analyses: the potential to affect the binding of transcription factors in Regulomedb(34), Genome Wide Annotation of VARIants(GWAVA), Genomic Evolutionary Rate Profiling (GERP)(35), brain-tissue-specific allelic effect on gene-expression(eQTL analysis) based on GTEX and BRAINEAC, and brain-tissue-specific allelic effect on DNA methylation in CpG loci(meQTL analysis). Details of these tools and analyses are described in the Supplement.

Replication analysis

Regional heritability in the significant block identified in GS:SFHS

Individual genotypes in UKbiobank and PGC2-MDD(22 cohorts) were used to estimate the regional heritability of the target haplotype block in the two samples. The two-GRM model (rGRM+cGRM) was applied to provide accurate estimates. For PGC2-MDD, the regional heritability was estimated for each of the 7 groups defined based on country of ancestor (Table s1) as well as for the combined dataset.

Single-SNP based association test for the five significant SNPs(genotyped) within the significant block identified in GS:SFHS

For UK Biobank the single-SNP-based association tests were performed using a logistic model in PLINK(36). Covariates included age, sex, centre, batch and 15 principal components provided by UK Biobank. For PGC2-MDD the association test was performed using a logistic model for each individual cohort. Covariates include sex and 20 principal components (the age variable was not yet available for the full dataset at the time of this study). Meta-analysis was performed across all cohorts in each group to generate group-level association statistics. The meta-analysis was performed using the ‘metagen’ function in R package ‘meta’. For the DS sample the GWAS summary statistics were downloaded from the website of social science GWAS consortium (<http://www.thessgac.org/#!/data/kuzq8>).

Results

Genome-wide Haplotype-block-based regional Heritability Mapping (HRHM) was carried out for 49,637 haplotype-blocks using 56,1125 genotyped common SNPs in GS:SFHS for MDD ($N_{\text{case}}=2,659$, $N_{\text{control}}=17,237$). The regional heritability from each haplotype-block was tested using a pre-adjusted-GRM strategy in the linear mixed model. The Manhattan plot and qqplot for the LRT are shown in Figure 1. One haplotype-block covering a 24kb region in the intron region and a proportion of an adjacent exon of gene *TOX2* exceeded the genome-wide significant threshold ($P_{\text{Bonf_threshold}}=1.01 \times 10^{-6}$): hg19:chromosome20:42555671-42579473 ($P_{\text{LRT}}=8.86 \times 10^{-7}$)

(Figure 1). The two-GRM model confirmed the significance of this haplotype-block ($P_{lrt}=5.6 \times 10^{-7}$) and the regional heritability (h_g^2) was estimated to be 0.008(0.006). The regional heritability of this block was more significant in females MDD ($h_g^2=0.009$, $se=0.007$, $P_{lrt}=5.64 \times 10^{-5}$, $N_{case}=1,893$, $N_{control}=9,818$) than in males MDD ($h_g^2=0.003$, $se=0.004$, $P_{lrt}=0.02$, $N_{case}=765$, $N_{control}=7,420$).

We further performed a series of association tests to disentangle the signal detected by HRHM in the significant block. Using the single-SNP-based association test, five of the nine genotyped common SNPs within the hit block were significantly associated with MDD (Table 1, s2). The five significant SNPs were in high linkage disequilibrium (LD) with each other (Figure 1D) and their minor alleles showed a consistent negative effect on the risk of MDD with the odds ratio ranging from 0.785 to 0.833 (Table 1). Haplotype-based association tests for haplotypes derived from the nine SNPs showed that two out of the seven common haplotypes (frequency ≥ 0.01) were associated with MDD. One of these haplotypes contains the minor (protective) alleles of the five single-SNP-level significant SNPs and one contains the major (risk) alleles. The size and the direction of the effects of the two haplotypes were consistent with that estimated from the single-SNP based tests (odds ratio: 0.792 for the protective haplotype and 1.232 for the risk haplotype) (Table 2). Additional association tests on sub datasets (unrelated and case-parents-trio) showed that the risk haplotype was significantly associated with MDD in the unrelated dataset (Table s3), whereas the protective haplotype was significant in case-parents-trio dataset (Table s4).

The significant block overlapped with an enhancer active in multi-tissues and cell lines including astrocytes (Figure 2A)(37), and multiple alternative transcription start sites (TSSs) including a TSS primarily expressed in thalamus (the TSS labeled as ‘p3@TOX2’ in Figure 2A)(37), suggesting a potential regulatory role. To link the association signal from single variants with the potentially functional effects of those variants on disease-relevant biological processes, we identified 38 imputed SNPs in the target block significantly associated with MDD (Table s5), and predicted their potentially regulatory function using multiple predictors and statistics of non-coding DNA function, including the likelihood of affecting transcription factor binding, multi-genome-wide properties, evolutionary conservation, the *cis*-effect on gene expression of genes within a distance of 1MB and on DNA methylation. Among the 38 SNPs, two of them were annotated to be ‘likely to affect TF binding’ (score=2b) by regulomeDB, five obtained a GWAVA-TSS score ≥ 0.5 (suggesting ‘functional’) and five obtained a GERP-score >2 (suggesting ‘constrained’) (Table s6). Tissue-specific snp-*cis*-gene-expression (*cis*-eQTL) analyses were performed for the 38 SNPs using 11 brain tissues from GTEX and 10 brain tissues from BRAINEAC. The results from GTEX showed that the genotype of 30 of the 38 SNPs significantly stratify the expression of gene RP1-269M15.3 (LncRNA) in the tissue ‘Nucleus accumbens basal ganglia’ with the minor alleles significantly up-regulating the RNA expression level (Table s7)(Figure 2B). The results from BRAINEAC suggested that all of the 38 SNPs significantly stratify the expression of gene *TOX2* in frontal cortex (minor allele induces up-regulation) (Figure 2C) and gene C20orf62 (LncRNA) (minor allele induces down-regulation) in Cerebellar cortex (Table s8,s9). The results from meQTL analysis suggested that 30 of the 38 SNPs are significant meQTL SNPs in frontal cortex, and that particularly 19 of them significantly stratify DNA methylation of a

CpG locus cg24403644 (minor allele induces hypo-methylation) (Table s10). cg24403644 is located in a cluster of TSSs in *TOX2* (Figure 2) and shows differential methylation between human fetal and postnatal lifetime in frontal cortex and during the fetal brain development(38,39). Among significant SNPs in the *cis*-eQTL and *cis*-meQTL analyses, rs79645278 located in the peak of active enhancer (in astrocytes and other cell lines), and was predicted to be ‘likely to affect TF binding’ (2b) in RegulomeDB, having a GWAVA-TSS score of 0.5 and a GERP-score of 2.31(Figure 2A,B,C, Table s6).

The regional heritability detected in the hit block was replicated in the UK-Ireland group in PGC2-MDD with nominal significance ($P_{irr}=0.049$, $h_g^2=0.001$, $se=0.001$), whilst it was not significant in other groups in PGC2-MDD and UK Biobank (Table s11). The single-SNP-based association test for the five significant SNPs (genotyped) in this block identified in GS:SFHS showed that all five were replicated in the DS sample; all five were also replicated in the UK-Ireland group in PGC2-MDD (Table 1. Results for individual cohorts were shown in Table s12 and Figure s1), but not in other PGC2-MDD groups or in the meta-analysed combined PGC2-MDD sample (Table s13); none of the five SNPs were replicated in the UK Biobank sample but all showed the same consistent direction of effect with that reported in the discovery sample (Table 1, Figure s1). Meta-analysis using all independent UK-Ireland replication samples (UK Biobank + 4 cohorts in PGC2-MDD-UK_Ireland) showed that all of the five SNPs reached nominal significance (Table s13, consistent sign with GS:SFHS as shown in Figure 3, using SNP rs6093898 as an example.

Discussion

The current study used a combination of genome-wide haplotype-block-based regional heritability mapping (HRHM), localized association tests and functional prediction to identify candidate genomic region associated with MDD. Using a large Scottish cohort GS:SFHS, a genome-wide significant haplotype block located in gene *TOX2* was identified by HRHM as a risk region for MDD. Association tests using both single SNPs and haplotypes within this block highlighted candidate contributing genetic variants for MDD. Replication analyses showed that the regional heritability in this block was nominally significant in the UK-Ireland groups in PGC2-MDD. The SNP-level association signals within the hit block were replicated in the UK-Ireland group in PGC2-MDD and a study of depressive symptom (DS) which has overlapping subjects from PGC2-MDD and UK Biobank.

As shown in this study, compared with single-SNP-based genome-wide association methods (GWAS), HRHM provided following advantages: (1) a smaller number of tests were performed therefore a less stringent threshold of genome-wide significance was applied. (2) Haplotype blocks rather than single SNPs were the unit of mapping, these are therefore relatively less dependent on the density of the genotype arrays and do not require the same SNPs to be typed or imputed in replication study. (3) HRHM applied a linear mixed model accounting for both polygenic component and family structure, and can be applied to both population and family data. (4) Since haplotype blocks were used as the unit of mapping, the identified locus has a less complex LD structure (Figure 1D), which will benefit the downstream identification of candidate variants.

To date, published GWASs have mapped associated variants to very few genes for MDD (*LHPP*, *SIRT1*, *TMEM161B-MEF2C* and *NEGR1*)(5,7). In this study, the identified haplotype block was located in gene *TOX2* (TOX high mobility group box family member 2, also known as *GCXI*), indicating a new candidate gene for MDD. *TOX2* is a putative transcriptional activator involved in the hypothalamo-pituitary-gonadal system(40), and is located in a large genomic region which has been previously reported as associated with depression symptoms in psychotic illness(41,42). The same locus has also been weakly-associated with conduct disorder in a previous study(43). Using available databases, we found that convergent evidence from TSS by Fantom5 annotation (Figure 2A), histone modification markers and Dnase peaks representing active enhancers by ENCODE annotation (Figure 2A), and transcription factor binding prediction by RegulomeDB (Table s6) suggested a regulatory function of this block. To test for the potential effects of the variants within the block on gene expression, we performed brain-tissue specific *cis*-QTL analysis for SNPs significantly associated with MDD within the block. The expression of a LncRNA RP1-269M15.3 was significantly up-regulated by the minor alleles (minor alleles are protective to MDD as shown in Table 1, s5) of candidate SNPs within the block in Nucleus Accumbens, a tissue having been previously implicated in MDD(44). RP1-269M15.3 was a multi-exon LncRNA with a multi-species conserved region (Figure s2A) and was only expressed specifically in brain tissues (Figure s2B) and therefore of potential function in brain tissues. Similarly, the expression of gene *TOX2* was significantly up-regulated by the minor alleles of candidate SNPs in frontal cortex, a relevant tissue of MDD as well (45). The regulatory effect of MDD-associated SNPs in gene *TOX2* in frontal cortex is further supported by the meQTL

analysis on the same tissue. Combined with the fact that all of the 19 SNPs are both meQTL and eQTL SNPs for gene *TOX2* in frontal cortex and the fact that hypo-methylation has been previously suggested to be correlated with up-regulation of gene expression(46), consistent evidence from both methylation and gene expression data indicated that the minor alleles (protective) of MDD-associated SNPs up-regulate the gene expression of *TOX2* in frontal cortex(Table s8,s10). Interestingly, the brain-specific expression of both RP1-269M15.3 and *TOX2* were highly correlated ($r \geq 0.7$) with a number of depression-related genes (for example, *LRFN5*, *GRM7*, *CRH*)(47,48) in brain development (<http://brainspan.org>) (Table s14, s15), suggesting that the expression networks involving those genes were potential targets of the effects from candidate variants. These results are consistent with a previous study suggesting an overrepresentation of MDD GWAS significant loci in CNS expression and the regulation of gene expression in CNS during development(7).

The regional heritability in the identified block was nominally significant only in the UK-Ireland group of PGC2-MDD. The five significant genotyped SNPs within the block identified in GS:SFHS were replicated in DS sample and in the UK-Ireland group in PGC2-MDD. UK Biobank sample failed to replicate any of them, although they showed consistent sign of effect. Those results are likely attributable to the phenotyping differences (diagnosed MDD in GS:SFHS, mostly diagnosed MDD in PGC(49), putative MDD in UK Biobank and depressive symptom in DS), and the clinical heterogeneity within MDD across PGC2-MDD groups as shown in Table s10)(12). Notably, UK-Ireland which shows the most consistent replication results are from the same country/region with GS:SFHS, so its cohorts are likely to have a

similar local genomic recombination pattern and LD structure with GS:SFHS and potentially carry alleles not common in other European cohorts, which may explain the better replication result from this group (Figure 3,s1).

There are, however, several limitations in the current study. Firstly, the re-adjustment strategy applied to genome-wide HRHM, whilst it reduced the computational burden, it was potentially excessively conservative in reporting true associations (observed LRT statistics were depleted from expectation, as shown in Figure 1D), which consequently reduced the power of HRHM(50). Secondly, phenotypic difference among discovery and replication samples impeded the complete replication of findings across all samples. UK Biobank samples are also from the same country/region as GS:SFHS, as are the UK-Ireland group of PGC2-MDD, but currently UK Biobank only have putative MDD information available for a small subset of genotyped participants. Ongoing clinical assessment of MDD and the genotyping work on this sample will potentially provide more power to the replication analysis for our findings in future data releases.

Conclusion

The present study showed the first application of genome-wide HRHM to a psychiatric disorder. A genome-wide significant region was identified by HRHM and the contributing genetic effect was localized to variants and haplotypes within the block. The results were partly replicated in two independent samples. Functional prediction and *cis*-eQTL analyses suggested that the genotype of associated-variants

within the block stratified the gene expression of a potentially functional LncRNA RP1-269M15.3 and gene *TOX2* in MDD-relevant brain tissues, which should be explored in further studies.

Data Access:

GS:SFHS data is available to researchers on application to the Generation Scotland Access Committee (access: <http://generationscotland.org>). The managed access process ensures that approval is granted only to research which comes under the terms of participant consent.

Financial Disclosures

AMMcI has previously received grant support from Pfizer, Lilly and Janssen. These studies are not connected to the current investigation. The author YNZ acknowledge the support from China Scholarship Council. The authors TKC and AMMcI acknowledge with gratitude the financial support received for this work from the Dr Mortimer and Theresa Sackler Foundation. PAT, DJP, IJD and AMMcI are members of The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. DJM is an NRS Fellow, funded by the CSO. PN and CSH acknowledge support from the MRC. All other authors report no biomedical financial interests or potential conflicts of interest.

Acknowledgements

We are grateful to the families who took part in GS:SFHS, the GPs and Scottish School of Primary Care for their help in recruiting them, and the whole GS team, which includes academic researchers, clinic staff, laboratory technicians, clerical workers, IT staff, statisticians and research managers. This work is supported by the Wellcome Trust through a Strategic Award, reference 104036/Z/14/Z. The Chief Scientist Office of the Scottish Government and the Scottish Funding Council provided core support for Generation Scotland. GS:SFHS was funded by a grant from the Scottish Government Health Department, Chief Scientist Office, number CZD/16/6.

Membership List of Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (PGC)

Stephan Ripke, Naomi R. Wray, Cathryn M. Lewis, Steven P. Hamilton, Myrna M. Weissman, Gerome Breen, Enda M. Byrne, Douglas H.R. Blackwood, Dorret I. Boomsma, Sven Cichon, Andrew C. Heath, Florian Holsboer, Susanne Lucae, Pamela A.F. Madden, Nicholas G. Martin, Peter McGuffin, Pierandrea Muglia, Markus M. Nothen, Brenda P. Penninx, Michele L. Pergadia, James B. Potash, Marcella Rietschel, Danyu Lin, Bertram Müller-Myhsok, Jianxin Shi, Stacy Steinberg, Hans J. Grabe, Paul Lichtenstein, Patrik Magnusson, Roy H. Perlis, Martin Preisig, Jordan W. Smoller, Kari Stefansson, Rudolf Uher, Zoltan Kutalik, Katherine E. Tansey, Alexander Teumer, Alexander Viktorin, Michael R. Barnes, Thomas Bettecken, Elisabeth B. Binder, René Breuer, Victor M. Castro, Susanne E. Churchill,

William H. Coryell, Nick Craddock, Ian W. Craig, Darina Czamara, Eco J. De Geus, Franziska Degenhardt, Anne E. Farmer, Maurizio Fava, Margarita Rivera, Josef Frank, Vivian S. Gainer, Patience J. Gallagher, Scott D. Gordon, Sergey Goryachev, Magdalena Gross, Michel Guipponi, Anjali K. Henders, Bernhard T. Baune, Stefan Herms, Ian B. Hickie, Susanne Hoefels, Witte Hoogendijk, Jouke Jan Hottenga, Dan V. Iosifescu, Marcus Ising, Ian Jones, Lisa Jones, Tzeng Jung-Ying, James A. Knowles, Isaac S. Kohane, Martin A. Kohli, Ania Korszun, Mikael Landen, William B. Lawson, Glyn Lewis, Donald MacIntyre, Wolfgang Maier, Manuel Mattheisen, Patrick J. McGrath, Andrew McIntosh, Alan McLean, Christel M. Middeldorp, Lefkos Middleton, Stefan Kloiber, Grant M. Montgomery, Shawn N. Murphy, Matthias Nauck, Willem A. Nolen, Dale R. Nyholt, Michael O'Donovan, Högni Oskarsson, Nancy Pedersen, William A. Scheftner, Andrea Schulz, Thomas G. Schulze, Stanley I. Shyn, Engilbert Sigurdsson, Susan L. Slager, Johannes H. Smit, Hreinn Stefansson, Michael Steffens, Thorgeir Thorgeirsson, Federica Tozzi, Jens Treutlein, Manfred Uhr, Edwin J.C.G. van den Oord, Gerard Van Grootheest, Henry Völzke, Jeffrey B. Weilburg, Gonneke Willemsen, Frans G. Zitman, Benjamin Neale, Mark Daly, Douglas F. Levinson, Patrick F. Sullivan

Legends

Figure 1 Genome-wide haplotype-block-based regional heritability mapping(HRHM) results on MDD in GS:SFHS. A: Manhattan plot. Each point represents a haplotype block. The location of the point is the mid-position of the haplotype block. B: qqplot for the LRT. The LRT statistics distributed as a mixture of 0 and chi-squared (df=1) distribution. C: zoom in region of the hit haplotype block region in Chromosome 20. D: LD structure within the hit haplotype block in GS:SFHS. The block is located in gene *TOX2*, it contains nine genotyped common SNPs(blue box) and five of them are in high LD(red arrow) in GS:SFHS.

Figure 2 Functional prediction of the hit haplotype block. A: functional annotation of the hit block. The hit haplotype block (red bar on the left top showing the block and blue bars showing the genotype SNPs in GS:SFHS) is located in the intron region and a proportion of an adjacent exon of gene *TOX2*, overlapped with Fantom5 enhancers and Transcription start sites, and regulatory-relevant histone modification peaks(H3K27Ac and H3K4Me1). Within the block, 38 imputed SNPs were associated with MDD, using SNP rs79645278(pink) as an example. This SNP is located in the peak of active enhancer in astrocyte(highlighted with blue line). B,C: boxplots showing tissue specific effect from SNPs that are both associated with MDD in GS:SFHS and gene expression, using SNP rs79645278 as an example. B: the minor allele of rs79645278 up-regulates the expression of a LncRNA RP1-269M15.3 in the tissue ‘Nucleus accumbens basal ganglia’. C: the minor allele of rs79645278 up-regulates the expression of gene *TOX2* in Frontal cortex. Abbreviations: cerebellar cortex (CRBL), frontal cortex (FCTX), hippocampus (HIPPO), medulla (specifically inferior olivary nucleus, MEDU), occipital cortex (specifically primary visual cortex,

OCTX), putamen (PUTM), substantia nigra (SNIG), thalamus (THAL), temporal cortex (TCTX) and intralobular white matter (WHMT).

Figure 3 Forest plot showing meta-analysis for single-SNP-based association test on GS:SFHS and all UK/Ireland replication samples(four PGC2-MDD cohorts and UK biobank), using SNP rs6093898 as an example.

Accepted manuscript

Reference

1. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, et al. (2013) Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med* 10: e1001547.
2. Sullivan PF, Neale MC, Kendler KS (2000) Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 157: 1552-1562.
3. Lohoff FW (2010) Overview of the genetics of major depressive disorder. *Curr Psychiatry Rep* 12: 539-546.
4. Schizophrenia Working Group of the Psychiatric Genomics C (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511: 421-427.
5. consortium C (2015) Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* 523: 588-591.
6. Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, et al. (2013) A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 18: 497-511.
7. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. (2016) Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet* advance online publication.
8. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* 106: 9362-9367.
9. Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM (2015) Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. *PLoS Genet* 11: e1004969.
10. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, et al. (2011) Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med* 9: 90.
11. Flint J, Kendler KS (2014) The genetics of major depression. *Neuron* 81: 484-503.
12. Milaneschi Y, Lamers F, Peyrot WJ, Abdellaoui A, Willemsen G, Hottenga JJ, et al. (2015) Polygenic dissection of major depression clinical heterogeneity. *Mol Psychiatry*.
13. Wray NR, Maier R (2014) Genetic Basis of Complex Genetic Disease: The Contribution of Disease Heterogeneity to Missing Heritability. *Current Epidemiology Reports* 1: 220-227.
14. Nagamine Y, Pong-Wong R, Navarro P, Vitart V, Hayward C, Rudan I, et al. (2012) Localising loci underlying complex trait variation using Regional Genomic Relationship Mapping. *PLoS One* 7: e46501.
15. Uemoto Y, Pong-Wong R, Navarro P, Vitart V, Hayward C, Wilson JF, et al. (2013) The power of regional heritability analysis for rare and common variant detection: simulations and application to eye biometrical traits. *Front Genet* 4: 232.
16. Riggio V, Matika O, Pong-Wong R, Stear MJ, Bishop SC (2013) Genome-wide

- association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. *Heredity (Edinb)* 110: 420-429.
17. Shirali M, Pong-Wong R, Navarro P, Knott S, Hayward C, Vitart V, et al. (2016) Regional heritability mapping method helps explain missing heritability of blood lipid traits in isolated populations. *Heredity (Edinb)* 116: 333-338.
18. Shirali M. Using Haplotype Mapping to Uncover the Missing Heritability: A Simulation Study; 2014. Asas.
19. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. (2006) Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genet* 7: 74.
20. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. (2013) Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *International Journal of Epidemiology* 42: 689-700.
21. First MB, Spitzer RL, Gibbon M, Williams JB (2001) Structured clinical interview for DSM-IV-TR axis I disorders—non-patient edition. New York State Psychiatric Institute, New York.
22. Fernandez-Pujals AM, Adams MJ, Thomson P, McKechnie AG, Blackwood DH, Smith BH, et al. (2015) Epidemiology and Heritability of Major Depressive Disorder, Stratified by Age of Onset, Sex, and Illness Course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* 10: e0142197.
23. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 12: e1001779.
24. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. (2013) Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS One* 8: e75362.
25. Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, et al. (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* 45: 984-+.
26. Zeng Y, Navarro P, Fernandez-Pujals AM, Hall LS, Clarke T-K, Thomson PA, et al. (2016) A Combined Pathway and Regional Heritability Analysis Indicates NETRIN1 Pathway is Associated with Major Depressive Disorder. *Biological Psychiatry*.
27. Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG, Fontana MA, et al. (2016) Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet*.
28. Cebamanos L, Gray A, Stewart I, Tenesa A (2014) Regional heritability advanced complex trait analysis for GPU and traditional parallel architectures. *Bioinformatics* 30: 1177-1179.
29. Visscher PM, Hemani G, Vinkhuyzen AA, Chen GB, Lee SH, Wray NR, et al. (2014) Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet* 10: e1004269.

30. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88: 76-82.
31. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, et al. (2013) Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. *Plos Genetics* 9.
32. Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T, Leo P, et al. (2013) Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nature Genetics* 45: 730-+.
33. Wang M, Lin SL (2014) FamLBL: detecting rare haplotype disease association based on common SNPs using case-parent triads. *Bioinformatics* 30: 2611-2618.
34. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 22: 1790-1797.
35. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S (2010) Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP plus. *Plos Computational Biology* 6.
36. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575.
37. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. (2015) Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol* 16: 22.
38. Jaffe AE, Gao Y, Deep-Soboslay A, Tao R, Hyde TM, Weinberger DR, et al. (2016) Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat Neurosci* 19: 40-47.
39. Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CC, O'Donovan MC, et al. (2015) Methylomic trajectories across human fetal brain development. *Genome Res* 25: 338-352.
40. Kajitani T, Mizutani T, Yamada K, Yazawa T, Sekiguchi T, Yoshino M, et al. (2004) Cloning and characterization of granulosa cell High-Mobility group (HMG)-box-protein-1, a novel HMG-box transcriptional regulator strongly expressed in rat ovarian granulosa cells. *Endocrinology* 145: 2307-2318.
41. Zhang XY, Bigdeli TB, Maher BS, Zhao Z, van den Oord EJCG, Thielson DL, et al. (2011) Comprehensive Gene-Based Association Study of a Chromosome 20 Linked Region Implicates Novel Risk Loci for Depressive Symptoms in Psychotic Illness. *PLoS ONE* 6: e21440.
42. Fanous AH, Neale MC, Webb BT, Straub RE, O'Neill FA, Walsh D, et al. (2008) Novel linkage to chromosome 20p using latent classes of psychotic illness in 270 Irish high-density families. *Biological Psychiatry* 64: 121-127.
43. Dick DM, Aliev F, Krueger RF, Edwards A, Agrawal A, Lynskey M, et al. (2011) Genome-wide association study of conduct disorder symptomatology. *Mol Psychiatry* 16: 800-808.
44. Pizzagalli DA, Holmes AJ, Dillon DG, Goetz EL, Birk JL, Bogdan R, et al. (2009) Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. *Am J Psychiatry* 166: 702-710.
45. Shelton RC, Claiborne J, Sidoryk-Wegrzynowicz M, Reddy R, Aschner M, Lewis

- DA, et al. (2011) Altered expression of genes involved in inflammation and apoptosis in frontal cortex in major depression. *Molecular Psychiatry* 16: 751-762.
46. Jones PA, Laird PW (1999) Cancer epigenetics comes of age. *Nature Genetics* 21: 163-167.
47. Nho K, Ramanan VK, Horgusluoglu E, Kim S, Inlow MH, Risacher SL, et al. (2015) Comprehensive Gene- and Pathway-Based Analysis of Depressive Symptoms in Older Adults. *Journal of Alzheimers Disease* 45: 1197-1206.
48. Holsboer F, Ising M (2008) Central CRH system in depression and anxiety--evidence from clinical studies with CRH1 receptor antagonists. *Eur J Pharmacol* 583: 350-357.
49. Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G, et al. (2012) A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry* 18: 497-511.
50. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL (2014) Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics* 46: 100-106.

SNP information					Discovery: GS:SFHS				Replication1:UK Biobank				Replication2:PGC2-MDD (UK Ireland)				Replication3:D S		
rs id	Ch	Pos	A1	A2	OR	log OR	se(log OR)	P	OR	log OR	se(log OR)	P	OR	log OR	se(log OR)	P	Beta	se	P
rs6017218	20	42555737	G (C)	T (A)	0.833	-0.183	0.041	2.44E-04	0.947	-0.055	0.030	0.068	0.842	-0.172	0.068	0.011	-0.013	0.005	0.007
rs6031242	20	42556096	G (C)	A (T)	0.832	-0.184	0.043	4.36E-04	0.948	-0.054	0.032	0.090	0.859	-0.153	0.071	0.032	-0.012	0.005	0.018
rs6031245	20	42559531	T (A)	C (G)	0.783	-0.244	0.045	2.30E-05	0.958	-0.043	0.035	0.225	0.843	-0.171	0.076	0.024	-0.015	0.006	0.011
rs6093898	20	42566577	G (C)	A (T)	0.783	-0.245	0.045	2.03E-05	0.958	-0.043	0.035	0.222	0.848	-0.165	0.075	0.028	-0.016	0.006	0.006
rs4812767	20	42568829	T (A)	C (G)	0.785	-0.242	0.045	2.57E-05	0.961	-0.040	0.035	0.253	0.840	-0.174	0.075	0.021	-0.016	0.006	0.006

Table 1. Single-SNP-based association test results for five MDD-associated SNPs in discovery and replication samples.

Haplotype	Frequency	Beta(linear)	se(Beta(linear))	OR	logOR	se(logOR)	P	Adjusted P
TAGCGACCT	0.120	0.026	0.005	1.232	0.209	0.058	2.47E-06	1.73E-05*
GGGTGGTCC	0.094	-0.024	0.006	0.792	-0.233	0.046	5.77E-05	4.04E-04*
TAGCAACT	0.118	-0.010	0.005	0.911	-0.093	0.045	6.10E-02	4.27E-01
TAGCGACTC	0.311	0.006	0.004	1.052	0.051	0.035	1.24E-01	8.71E-01
GAGCAACT	0.012	-0.012	0.016	0.897	-0.109	0.131	4.60E-01	1.00E+00
TAGCAACC	0.015	-0.010	0.014	0.916	-0.088	0.120	5.05E-01	1.00E+00
TATCGACTC	0.304	-0.002	0.004	0.980	-0.020	0.033	5.59E-01	1.00E+00

Table 2 Haplotype-based association test results for common haplotypes derived from the nine genotyped common SNPs in GS:SFHS.

Adjusted P: Bonferroni method adjusted P values.

